



The European Future Technologies Conference and Exhibition 2011

## eMorph: Towards Neuromorphic Robotic Vision

Chiara Bartolozzi<sup>a</sup>, Charles Clercq<sup>a</sup>, Neeraj Mandloi<sup>a</sup>, Francesco Rea<sup>a</sup>,  
Giacomo Indiveri<sup>b</sup>, Daniel Fasnacht<sup>b</sup>, Giorgio Metta<sup>c</sup>, Michael Hofstätter<sup>d</sup>,  
Ryad Benosman<sup>e</sup>

<sup>a</sup> Italian Institute of Technology, via Morego 30, Genova, Italy

<sup>b</sup> Institute of Neuroinformatics (ETH—UZH), Winterthurerstrasse 190, Zurich, Switzerland

<sup>c</sup> Università degli Studi di Genova, viale Causa 13, Genova, Italy

<sup>d</sup> Austrian Institute of Technology, Donau-City-Strasse 1, Vienna, Austria

<sup>e</sup> ISIR - Université Pierre et Marie Curie, Pyramide - Tour 55, BP 173, Paris, France

### Abstract

The eMorph project aims at introducing a new concept for vision in the field of humanoid robotics. The system that is currently being developed is inspired by the biology of mammalian visual systems, introducing concepts such as stimulus-driven signal acquisition and processing, together with space-variant sensor design coupled with active vision. This approach is leading to the realization of a system that goes beyond current thinking in robotic vision.

© conference organizers and published by Elsevier B.V. Open access under [CC BY-NC-ND license](#).

Selection and peer review under responsibility of FET11

**Keywords:** neuromorphic; humanoid robot; event-driven computation; vision

### 1. Introduction

State of the art computer vision has its roots on the concept of frames. In machine vision, the world is sampled at constant rate, leading to collections of static snapshots. The information within each frame and among frames is highly redundant and has very little dynamic content, making visual data transfer, storage and processing, time and resources consuming tasks. Tasks that seem effortless for animals such as object segmentation and motion evaluation (to cite a few) require full processing power by several CPUs; even then, the outcome is hardly robust nor computed with a speed that allows efficient interaction with the real world in real time. Understanding the principles behind the properties of biological systems gives the opportunity of engineering them in a new bio-inspired technology that will incorporate the strength of brain-like sensing and computing. This approach is crucial for allowing robots to robustly and reliably interact with the real-world in real-time and for getting closer to autonomous robotics. Brain computation is stimulus driven, triggered by significant events in the sensed world. It is designed to enhance variations and discontinuities, discarding redundancies at the lowest level. At the same time the relative encoding of the events, under the form of temporal and spatial contrast encoding, ensures a wide dynamic range; adaptation grants a form of automatic gain control that makes animals able to behave in a wide variety of different conditions. Within this context we are developing a vision system for the humanoid platform iCub (RobotCub project, <http://www.robotcub.org/>) that incorporates the principle of asynchronous event-driven sensing by using the asynchronous “Dynamic Vision Sensor” (DVS) [1] and developing a framework for event-driven computation: both at the hardware level, with a dedicated embedded digital processor, and at the software level, with the development of a paradigm for event-driven algorithms

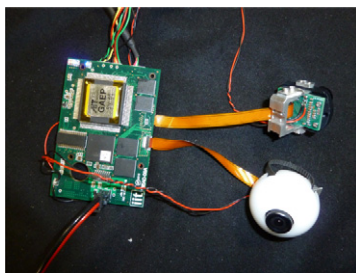


Figure 1. Picture of the miniaturized iHead board that hosts the GAEP and the FPGA, connected with the asynchronous eyes.

and procedures. At the same time, we are developing new vision sensors with higher complexity, with space-variant topology and the ability of extracting spatial derivatives and motion cues from the sensed environment.

## 2. Embodied asynchronous vision

The current implementation of the system comprises two asynchronous DVS sensors together with an embedded dedicated processor for asynchronous data processing, the General Address Event Processor (GAEP) [2], and a field programmable gate array (FPGA) connecting the sensors to the GAEP and then to a CPU. These custom hardware components are designed to communicate asynchronous data with a spike based “address-event representation” (AER) protocol [3]. A software module collects streams of asynchronous events that are then available for further processing and visualization. The asynchronous eyes replace the standard frame-based cameras inside the pupils of the iCub, while the standard cameras are placed on top of the robot’s head. This second visual input is used in parallel to the asynchronous cameras to complement the visual information with color-based images currently needed for some visual modules such as object-based attention, object segmentation and recognition.

### 2.1. The Vision Sensor

The asynchronous sensor mounted on the iCub is the well known DVS [1], an event-driven AER sensor that detects changes of contrast in its visual field and encodes them in a sequence of digital pulses (events) that convey information about the visual input in their relative timing. Each pixel independently responds to variations of contrast in the visual field, achieving a wide intra-scene dynamical range, that makes the sensor capable of responding to a large interval of contrast over different illuminations. The DVS is currently mounted on the iCub and used to develop procedures and algorithms for asynchronous vision. Concurrently, we are exploring alternatives for the design of vision sensors, which can optimize the trade-off between on-chip processing and resolution. We developed a monodimensional smart sensor capable of computing edge velocity and spatial derivative and that can report the location of the most interesting stimulus in the visual scene [4]. We also designed and sent to production an asynchronous sensor that allows the computation of both temporal and spatial contrast and that, in addition, has a space-variant arrangement of photoreceptors, achieving high resolution in the center (the fovea) and lower resolution in the periphery. This arrangement allows to increase sensor’s resolution, without increasing the number of pixels nor decreasing the camera field of view. This approach is helpful for reducing the data transfer, but it is feasible only for active vision systems, that allow the redirection of gaze towards regions of interest detected by an attentive system, that is the case of the humanoid platform iCub (Figure 1).

### 2.2. Embedded hardware

We developed a custom printed circuit board that hosts the GAEP, an FPGA and connectors to the asynchronous cameras and to the robot’s CPU. The FPGA merges and routes the events from the left and right sensors, streaming them to the GAEP and then from the GAEP to the robot’s CPU. Additionally, it hosts a dedicated serial line [3] that implements a fast asynchronous bus to connect post-processing neuromorphic chips that we plan to add in the next future. The general purpose address-event processor is based on a SPARC-compatible LEON3 core with a custom data interface for asynchronous sensor data. The GAEP responds to the need for transferring the inherently precise timing information of asynchronous events into the synchronous domain at high temporal resolution (submicrosecond),

combined with local processing capability functionality. In the final visual system it will perform pre-processing of visual data, conveying information about visual features (e.g. edge orientation) to the robot's core computing unit, where it will be used for extracting high-level information about the environment.

### 2.3. *Software modules*

The software procedures for handling and computing with events are built within YARP [5], the brain-inspired software infrastructure of the iCub, designed to implement high degree of parallelism by facilitating concurrent processing and distributed computing. In this context, we developed a module that collects events from the FPGA and broadcasts them to the YARP network, minimizing the overhead time that could come from multi-point communication between computing nodes. Any other module can read the events from the flux of data in the network. Among them, the *Frame Converter* module visualizes events as sequence of constant time frames. Events are read from the network and only those whose timestamps belong to the current frame are represented in their respective location of the visual field. While developing modules for event-based artificial vision, the representation of events as frames is a crucial step toward the innovative goal of integrating traditional visual perception with event-driven perception. The first step towards this goal is the integration of event-based attentional modules with the available iCub attentive system, needed to direct attention and gaze towards salient regions of the visual input. The YARP visual attention module is shaped on a model of stimulus-driven primates spatial attention [6], enriched with object-based components [7]. The current effort is the development of event-based feature maps based on motion, orientation, contrast, etc. that will be added to the computation of the current saliency map. At the same time we are developing algorithms and modules for event-based artificial vision, that can take advantage of the high temporal resolution and low data rate from asynchronous sensors, such as modules for the computation of optical flow.

## 3. Conclusions

We are developing a neuromorphic visual system for the humanoid platform iCub that will introduce a new concept of artificial vision in robotics. Our efforts span from the miniaturization and integration of dedicated hardware components of the system, resulting in a new asynchronous embedded module for the robot, to the design of new asynchronous space variant sensors, to end up with the development of a framework for asynchronous artificial vision. The implementation of such a system on an open source robotic platform capable of meaningful interaction with the real world in real time is the key to validate the advantages of the neuromorphic approach and make them available to a wide community.

## 4. Acknowledgments

This work was supported under the EU FP7 framework: ICT-STREP project eMorph, 231467 on Embodied Intelligence.

## References

- [1] P. Lichtsteiner, C. Posch, T. Delbruck, An  $128 \times 128$  120 dB 15  $\mu$ s latency temporal contrast vision sensor, *IEEE J. Solid State Circuits* 43 (2) (2008) 566–576.
- [2] M. Hofstätter, P. Schön, C. Posch, A sparc-compatible general purpose address-event processor with 20-bit 10 ns-resolution asynchronous sensor data interface in 0.18  $\mu$ m cmos, in: *International Symposium on Circuits and Systems, ISCAS 2010, IEEE, 2010*, pp. 4229–4232, doi:10.1109/ISCAS.2010.5537575.
- [3] D. Fasnacht, A. Whatley, G. Indiveri, A serial communication infrastructure for multi-chip address event system, in: *International Symposium on Circuits and Systems, ISCAS 2008, IEEE, 2008*, pp. 648–651, doi:http://dx.doi.org/10.1109/ISCAS. 2008.4541501. URL [http://ncs.ethz.ch/pubs/pdf/Fasnacht\\_et al08.pdf](http://ncs.ethz.ch/pubs/pdf/Fasnacht_et al08.pdf).
- [4] C. Bartolozzi, N.K. Mandloi, G. Indiveri, Attentive motion sensor for mobile robotic applications, in: *International Symposium on Circuits and Systems, ISCAS 2011, IEEE, 2011*.
- [5] P. Fitzpatrick, G. Metta, L. Natale, Towards long-lived robot genes, *Robotics and Autonomous Systems* 56 (1) (2008) 29–34, doi:10.1016/j.robot.2007.09.014.
- [6] L. Itti, C. Koch, Computational modeling of visual attention, *Nature Reviews Neuroscience* 2 (3) (2001) 194–203.
- [7] F. Orabona, G. Metta, G. Sandini, Object-based visual attention: a model for a behaving robot, in: *Computer Vision and Pattern Recognition – Workshops, 2005. IEEE Computer Society Conference on CVPR Workshops, 2005*, p. 89, doi:10.1109/CVPR. 2005.502.